Course Name: Technical Communication for Engineers

Professor Name: Prof. Arun K. Saraf

Department Name: Department of Earth Sciences

Institute Name: Indian Institute of Technology Roorkee

Week- 11

Lecture 11- Key components of a chart and types of plots

Hello everyone and welcome to discussion which is a continuation of our previous discussion which is about the key components of chart and types of plots. So we will continue on that line. In earlier discussion we had you know good and bad examples of you know bar charts and other things and their you know intricacies or problems associated with them. So in that continuation now I will bring the First good example of line graph as we have been discussing in the previous lecture that a lot of data can be simply represented using line graphs. And this is one of the examples like this we can show the you know two observations or two series of data we can plot and they can show very clearly where these are matching or where there are big mismatches are there. So, all the data can be plotted again you know some points are the same that all your axis should be marked should have a proper scale units. So, here like on this example we are having year on the x axis. number of words observed in y axis and since our values vary between 200 to 600 and therefore, here the 0 has not been kept because if we keep the 0 then what will happen below 200 everything will remain blank or empty and that does not look nice. So, it is very easy to change the scale, the origin of the scale, the origin value of the scale and then we get the plot. The purpose here that whatever the space which is available to us should be fully utilized and we should not waste or leave a large space empty.

So, as example I have given that if we keep these number of words that is by scale between 0 to 600 then the lower part of this line graph. will be completely blank and that does not look nice, does not give much impression, one. Secondly, the differentiation between these two, you know, number of words which are two different types which are mentioned here will not show that much of difference. So, it is always instead of starting 0, you can start with the nearest value or less than that value. Similarly, for the top value also, one should go for that level. And one should not like for example, why default somebody might plot between 0 to 1000. And then these two curves which are being shown here, two line graphs will squeeze further and no differentiation would be possible. So, therefore, it is a good example of a line graph. Now, we can see a bad example of line graph. The same point which I was saying here that this much area is

going empty and we are not achieving anything. Secondly, the differentiation or clarity which we had in the previous example is not as good as that one. Further, these lines parallel to x axis are also not required. So, these are the bad points about this line graph. Similarly, there are lot of other options like bar chart, line graph. Now, we can see the good example of a box plot in which we generally use these plots to show the error margins and other things. So, again here see that the beginning of y axis and end of y axis is between 5 to 25 not between 0 to 30 or 0 to 100 because we want to show here things much more clear. And, the second is the area which is available for plotting is almost fully utilized. So, that is the good part. And, axis are properly labled, marked. Their units are also written properly and that is the advantage. Plus, legend is also available that we are talking about this. and then G. magniro stris and then G. Scandens like this. So, a comparison can also be made between variables and the chart becomes very clear even to a layman. That is the advantage of you know applying a very simple logic, simple mind and you can really create impactful plots or graphs. the same data has been plotted where the y axis values have been taken between 0 to 100 and the large part of this plot is just underutilized or no utilized at all. And, the differentiation which was visible in the previous part like this one here are not coming at all here. So, then the, you know, the understanding which one would develop based on this bad example of box plot would not be appropriate and this is not good for such plotting. So, there is no problem.

We can get rid of the top part, rescaling the y axis and do the same thing. Similar kind of adjustments for y scale, instead of y scale, We can have also on X scale depending on the data which we are having or the series of data which we are having. And now a pie chart. Sometimes we also go for pie charts. Like here I am having example that rainfall by quarter.

So, we are having a title for that and different sections of this pie are showing different percentages and they are quarter 1, quarter 2, quarter 3, quarter 4 likewise. So, pie chart typically show one series of data and each data point show as a percentage of a whole. We can show as a percentage of whole or some other values but generally percentage is preferred in case of pie chart. And one can easily see the relative values in each other like if I start trying to understand this figure then I find that the quarter 4 had the maximum rainfall that is around 34.34 percent and whereas the quarter 1 that is in red colour had a rainfall of 15.47 percent. So, by seeing the size of pi, one gets the idea that in which quarter we had the maximum rain, in which quarter we had minimum rain and so on. So, a relative understanding about different data of the same series can be guessed or can be studied, can be understood very easily. So, again good example of pie charts. Even pie charts can be presented in form of grey scales or in black and white. Like here, pie chart has been presented and again shown in the percentage. And entire space whichever was available has been fully utilised. And all those things are there. And therefore, it becomes

a good example of pie chart. We can have a bad example of pie chart just to understand what kind of mistakes. First of all, there are too many, you know, pieces of pie and then interpretation of this becomes very difficult. There is no need of making this 3D thing. Because only part of this pie one is able to see, the remaining is not. And thirdly that since it is not a perfect circle, so it is making a biased you know opinion about a certain section or certain part of the pie. And of course, no labeling of percentage or any other thing is there. One thing is definitely there that is the legend but the values are not there.

So, we do not know which is representing how much percentage of the data. So, this we consider as a bad example of pie chart. Incomplete pie chart not giving information but because the 3D option was available, so one is trying to do that thing. Now, when we come to this scatter graphs, scatter graphs are plotted for to see the relation between two variables or correlation between two variables. So, here what you see that you know the two variables. So, one is the beak length and other is what is the wingspan This is the example from biology so I am taking from there that wings span what is the length of the wing. So, it is mentioned here and what is here mentioned is the beak length. So, is there any relationship between wing length and beak length and if we say if we try to fit a best fit line here then generally it will fit like this. So, it shows that yes, there is a linkage between this. Different species are also located at different locations and therefore, it is a good, we can consider as a good example of scattered plots because that is the purpose of two showing the relationship between two variables. Scatter graphs show x y data as we have seen and some of the interpretations very easily one can made between two or relationship between two data sets, two variables. Like in the first example we are having very strong correlation because they are just as we move along x axis the values are also of other variable or value other series is moving in the same way along y axis. Most of these dots are all along the best fit line and that best fit line is also having roughly 45 degree. So, there is a almost one to one correlation exist. The big positive correlation is there since this angle is anyway is in positive direction. So, but this is the scatter is much bigger than what along the best fit line. So, therefore it is declared as big positive relation. Similarly, for a negative relation it would be that as I move towards the x axis things are increasing but in case of the same data the things are decreasing and therefore it is a strong negative correlation. It is just opposite of example 1 and this is 3. So, this is just opposite of example 1. Similarly, we are having weak negative relation because the points are getting scattered all along that best fit line. Moderate negative correlation or sometimes we may get this kind of correlation that two parameters, two variables are not correlated at all. They are all randomly distributed. So, this gives a very good idea about the correlation or relationship between two data sets. So, what we do? We plot numerical values along these two axis and typically scatter graphs or plots are used to show the relationship between two sets of numeric values. It is also possible sometimes we require to see the relationship between three values. Therefore the three-dimensional plots or

bivariate plots or three-dimensional plots are also possible. But the angle in which it should be shown has to be chosen very carefully. Otherwise, the point which you would like to highlight may not come very clearly. So, this is the thing. For example, if we are having non-numeric values such as string categories or date, time datatype type then scatter plots should not be used at all. So, these are mainly used for numeric values if at all you are having. So, good example of scatter plot we have already seen in just 2-3 slides before and that scatter plots should be like this. The levels, units and thing and one more thing which one can add as per the requirements the best fit line and the equation also. This is the bad example of scatter plot because large area is underutilized.

And this relationship is just concentrated only in a small part. So, what should have been done that instead of having value between 0 to 90, the value should be chosen between 60 to 90. And similarly here, instead of value between 0 to 20, the value should have been chosen between 8 to say 18. And therefore, this thing between two datasets would have come very clearly, nicely and understandably. But here we are missing the opportunity to show the relationship or no relationship kind of thing. And same time we are also losing opportunities to utilize the full space. This you remember that whatever the space is available for display, we should fully utilize. This thing will also be true in case of PowerPoint slides also. So, do not take you know put everything on one corner or half of the slide rest remain empty. That is not a good practice. And that we will consider as bad example of a scatter plot or a slide. Now as just mentioned earlier that three-dimensional graphs are also possible. But as also mentioned that the angle in which they are shown should be very clear. Clear in sense that whatever I wanted to show should come very clearly to the audience in an understandable manner. So, here what we are seeing that an example of you know bar charts or three series are here and then so you would have then three axis x, y and z axis. So, this becomes my z axis, this is x axis, this is my y axis. So, in three dimensional or 3D graphs in which we can use area, bar, column, line, pie graphs all can be created. And, there are series along a third axis and like here the third axis is sales in units of a particular product is plotted there. So, these will give the perspective to use the and show the third dimension. So, it gives a third dimensional perspective about things.

But what is the problem here in this particular plot that the blue, you know, this stellar product is not being visible for this area and this area. We do not know what is happening. Secondly, this galactic is also not very clear. It seems like it is a continuous thing. So, I would consider a bad example of three-dimensional plot. So, three-dimensional plot only in special circumstances one should create one thing. And also you know the angle, the viewing angle from which direction you should that has to be chosen after lot of iterations, hit and try. And once you are assured that it is communicating everything whatever you wanted then it is good. And one of the good practices is just

always show to your colleagues, friends or students that whether they are able to understand this graph or not. Another way and nowadays you know the tools which have moved beyond our excel sheets or simple line graph or other things and that is it is becoming possible using GIS that is geographic information systems based.

And what is happening that now we are trying to attach each and every information with geographic co-ordinates. If it is possible to attach that information with geographic coordinates then we can plot in a truly on a map as example shown here. So, this is what it is showing the you know corona infected people like here of a certain date that confirm cases and then you are having active cases and other thing. So, one by one whichever you want to see a map that will come. And, here it is shown through the circles. So, the larger the size of the circle, more the infected people are there. So, the size is showing, location is showing and it is covering all the entire India. And, especially I have just brought a zoom part of that one and that is Uttarakhand was also shown here. So, you can see that once the data is plotted using geographic coordinates, then lot of patterns will come. And the questions will come and questions can be also answered and that is the advantage plotting using GIS. This was done very successfully during corona period of those two and half or three years. As you see here that you know the pattern is coming which part of India had the maximum infected people and which part or confirmed cases of corona which part did not have. For example, Rajasthan, And, the data shows that we were getting information from almost every district of Rajasthan. This is district level data. So, we are getting information from every district but the people infected are not as much as in like larger metros like Delhi or Mumbai or maybe Chennai and other parts here.

So, this is the advantage. Immediately one has to have little knowledge about geography, locations of different places and immediately you get the idea what is happening. See also that in central India again the cases were very less. But one can also correlate with the population density of the people. Because if population density like in these areas I know the population density is relatively low as compared to metro cities. So, where population was less, of course, the confirmed cases of corona were also less.

So, that was a complete understanding or correlation was established that where the density of the people or density of population is more, there are more chances of getting corona. Where people are living in a sparse area with less dense population density, there are less chances of getting corona. So, that kind of interpretation and the same thing it was also plotted for confirmed cases how they are moving with time, active cases and you know the cases which has recovered and that kind of thing. So, all kinds of information plus a bar chart is also there. So, just on a one screen, almost everything related with Uttarakhand was available in detail and related with all over the country that was also available. Like here, it says the total confirmed cases on that particular day were

like this. So, these kind of plots are also becoming very very useful. And you think that if instead of having a GIS based plots, if we would have just like bar chart or like these then they are not giving full information about the area. Because the location specific data is not here.But in case of GIS it is there. There were some other ways people were plotting and this is for the global data COVID-19 infection map and it was shown here with the number of deaths aggregated confirmed, active cases and or your graphs were also getting plotted. And we knew that it is the rise of confirmed cases was not in a linear fashion, it was complete in exponential phase. So, that is what you can also see. And once you want, these can be plotted on the map. In this particular case, it was you know mosaic of satellite image or Google earth that can be plotted. So, we are coming to the end of now this discussion of two lectures basically. So, this is the last part about this thing. So, when we are using power point charts can also be prepared like this. And this is the example of clustered column or clustered bar chart. where we are having 4 series of data that can be plotted and the relative you know variations can be shown very clearly.

Now finally, before I end, I would like to bring something new which I have recently learnt. So, I thought that I will also add here that is the plots to be avoided. This is very-2 important thing that pie charts are very bad way of displaying information. Though I have said that one of the ways of preparing graphs is pie chart. But people have mentioned and I will show the person's name also that he is Professor Carl Brumman, University of Wisconsin, Madison, USA. And he says that the eye is good at judging linear measurements and bad at judging relative areas like in pie charts. And bar charts or dot chart is a preferable way of displaying this type of data. So, therefore, I also feel that Professor Broman is quite correct that the psychology of humans especially while seeing things, you know, they prefer simple things. Our eyes and brain prefer simple things. So, comparatively pie chart is little complicated as line graphs or bar charts.

And therefore, only in certain circumstances, one should go for the pie charts. Otherwise, try to display your data as far as possible using simple charts, simple figures, simple diagrams and that is like line graphs or maybe bar charts. You have made pie charts and different slices are going here and there. And the one who is trying to judge or trying to understand the distribution of the data is not very comfortable as compared to if it would have been in a line or bar chart. So, with this I end this discussion. Thank you very much. Namaskar.